

# PREDIKSI PENENTUAN PENERIMA BEASISWA DENGAN METODE KNEAREST NEIGHBOURS

(Studi Kasus: Program Studi Sistem Informasi Fakultas Vokasi Universitas Airlangga)

Nasa Zata Dina<sup>1</sup>, Rachman Sinatriya Marjianto<sup>2</sup>

<sup>1,2</sup> Universitas Airlangga

Jl. Dharmawangsa Dalam Selatan 68, Surabaya, 60286, Indonesia

[nasazatadina@vokasi.unair.ac.id](mailto:nasazatadina@vokasi.unair.ac.id)

**Abstract**— The decision support system or prediction is expected to assist in determining scholarship recipients. This research is focused on case study of Information System Study Program of Faculty of Vocational, Universitas Airlangga. The Staf is overwhelmed because the large number of scholarship's applicants while the number of scholarship is limited. The scholarship quota itself is also provided for each study program based on the number of students in each study program. Because of the all-round administrative process is manual then the staf finds difficulties in determining the recipients of the scholarships that exist in each semester. As a result, some scholarships have not been targeted according to its designation. Sometimes, there are some students who receive double scholarships due to negligence in the determination of the scholarship recipients. Because of these problems, the decision support system or prediction in determining scholarship recipients is needed in order to build a computerized system in the environment of Information System Study Program Faculty of Vocational, Universitas Airlangga. Research method is used include data collection, system analysis, system design, system implementation and evaluation, and system testing. KNearest Neighbours Algorithm is used and the parameter value of K is 2. Poverty Index and Cumulative Achievement Index are the measured variable. The accuracy value that is generated by KNearest Neighbours algorithm with K = 2 is 75%.

**Keywords**— Classification, KNearest Neighbours, Scholarships, Vocational.

**Abstrak**— Sistem pendukung keputusan berupa prediksi penentuan penerima beasiswa online diharapkan dapat membantu dalam menentukan mahasiswa calon penerima beasiswa. Pada penelitian ini difokuskan pada studi kasus Program Studi Sistem Informasi Fakultas Vokasi, Universitas Airlangga. Banyaknya jumlah beasiswa yang masuk ke fakultas mengakibatkan staf kemahasiswaan kewalahan karena jumlah calon penerima beasiswa sangat banyak sedangkan beasiswa yang ditawarkan terbatas. Kuota beasiswa sendiri juga diberikan pada tiap program studi berdasarkan banyaknya jumlah mahasiswa pada tiap angkatan. Karena proses administrasi serba manual maka petugas kemahasiswaan kesulitan dalam menentukan penerima dari beasiswa yang ada pada tiap semesternya. Akibatnya beberapa beasiswa peruntukkannya belum tepat sasaran sehingga terkadang ada beberapa mahasiswa yang menerima beasiswa ganda akibat kelalaian dalam penentuan penerima beasiswa. Oleh karena permasalahan tersebut maka sistem pendukung berupa prediksi penentuan penerima beasiswa sangat dibutuhkan untuk saat ini demi membangun sistem terkomputerisasi yang terstruktur di lingkungan Program Studi Sistem Informasi Fakultas Vokasi, Universitas Airlangga. Metode penelitian yang digunakan mencakup pengumpulan data, analisis sistem, desain system, implementasi sistem dan evaluasi dan pengujian sistem. Sedangkan algoritma yang digunakan adalah Algoritma KNearest Neighbours. Nilai parameter K yang digunakan adalah 2. Indeks Kemiskinan dan Indeks Prestasi Kumulatif merupakan variabel yang diukur. Nilai akurasi yang dihasilkan oleh algoritma KNearest Neighbours dengan nilai K= 2 yaitu sebesar 75%.

**Keywords**— Beasiswa, Klasifikasi, KNearest Neighbours, Vokasi.

## I. PENDAHULUAN

Setiap universitas memiliki banyak skema bantuan pendidikan bagi mahasiswa yang kurang mampu namun dapat menunjukkan prestasi dalam bidang akademik. Banyaknya dana pendidikan yang diberikan pemerintah dan swasta kepada para mahasiswa membuat staf kemahasiswaan

kewalahan dan kebingungan dalam menyalurkan beasiswa kepada yang berhak. Selain jumlah beasiswa yang banyak tetapi juga jumlah mahasiswa yang akan menerima jumlahnya juga banyak. Pada tingkat fakultas saja sumber dana beasiswa cukup beragam, di antaranya dari DIKTI, pemerintah kota, perusahaan BUMN, UNAIR, yayasan serta beberapa perusahaan swasta. Tawaran beasiswa yang beragam tersebut

memperoleh tanggapan yang positif dari mahasiswa. Jumlah pelamar pun membludak sehingga staf kemahasiswaan cukup direpotkan dalam menentukan mahasiswa terpilih. Bahkan terkadang mahasiswa mendapatkan dua beasiswa sekaligus, bisa disebabkan karena kelalaian staf atau mahasiswa tidak melaporkan bahwa telah mendapat beasiswa. Salah satu cara untuk mengatasi kesulitan dalam penentuan penerima beasiswa adalah dengan dibantu sistem pendukung keputusan/prediksi. Waktu yang dibutuhkan lebih cepat dan hasilnya cukup efektif dalam menyimpan data penerima beasiswa.

Prediksi merupakan salah satu sistem pendukung keputusan membantu mengklasifikasi data testing ke dalam golongan-golongan tertentu sesuai dengan ketentuan awal atau ke dalam suatu kelas yang dilatih pada data training. Penggunaan algoritma dalam sistem pendukung keputusan merupakan hal yang mutlak karena terkait akurasi dan ketepatan rekomendasi atau prediksi. Algoritma klasifikasi yang cukup populer salah satunya adalah KNearest Neighbours. Algoritma KNearest Neighbours digunakan dengan menghitung bobot sesuai dengan jarak terdekat dari data testing ke data training sehingga dapat menggolongkan data berdasarkan jumlah kelas K [4].

Berdasarkan permasalahan yang telah dibahas sebelumnya maka dibutuhkan pendukung keputusan yang dapat membantu staf kemahasiswaan dalam menentukan penerima beasiswa. Sistem tersebut akan mengklasifikasikan data mahasiswa berdasarkan beberapa kriteria yang disepakati sesuai syarat dan ketentuan penerima beasiswa, sebagai contoh adalah Indeks Prestasi Kumulatif, kemampuan orang tua, dan sebagainya. Data training yang akan digunakan untuk mengklasifikasikan penerima diambil dari dataset penerima beasiswa sebelumnya sedangkan data testing diambilkan dari calon penerima beasiswa. Dari kedua data tersebut diperoleh dataset yang akurat sehingga pihak fakultas dapat mengatasi sejak awal permasalahan tersebut. Pendekatan algoritma yang digunakan adalah algoritma KNearest Neighbours karena algoritma tersebut cukup handal dalam melakukan prediksi dengan kedua variabel (IPK dan Penghasilan Orang Tua). Karena hasil akhir dari prediksi ini adalah “penerima” dan “bukan penerima” maka jumlah K yang dipakai adalah  $K=2$ . Sistem ini diharapkan mampu mempermudah kinerja staf bagian kemahasiswaan dalam menentukan penerima beasiswa yang paling sesuai di lingkungan fakultas vokasi. Selain itu tujuan dari penelitian ini adalah mengimplementasikan Algoritma KNearest Neighbours.

## II. TINJAUAN PUSTAKA

Pada bagian ini akan dijelaskan mengenai studi kepustakaan yang terkait dengan penelitian yang sudah dilakukan.

### A. Sistem Pendukung Keputusan

Sistem Pendukung Keputusan atau prediksi merupakan sistem berbasis pemecahan masalah yang menggunakan

algoritma kemudian akan diimplementasikan menjadi program komputer yang dapat digunakan oleh stakeholder. Masalah yang diselesaikan kebanyakan adalah masalah yang tidak memiliki struktur [5]. Sistem pendukung keputusan atau prediksi menggunakan data yang dikumpulkan lalu menampilkannya ke dalam antarmuka yang bisa dipahami oleh stakeholder kemudian diproses menggunakan algoritma dan menghasilkan sistem pendukung keputusan [6]. Sistem pendukung keputusan memiliki beberapa bagian di antaranya: (1) Manajemen Data yaitu data yang dikumpulkan disimpan ke dalam suatu database kemudian dikelompokkan berdasarkan kondisi dan syarat tertentu. Dalam pengelolaan datanya diatur oleh perangkat lunak khusus pengelolaan database yang disebut Database Management Systems (DBMS). (2) Manajemen Model digunakan untuk merepresentasikan suatu data atau bahkan hasil secara statistik sehingga dapat dianalisa data maupun hasilnya. (3) Komunikasi diperlukan pengguna agar dapat memberikan perintah maupun respon tertentu melalui antarmuka yang sudah dibuat. (4) *Knowledge Management* merupakan subsistem opsional yang dapat mendukung subsistem lain atau bertindak sebagai komponen yang berdiri sendiri.

### B. Proses Pengambilan Keputusan

Pengambilan keputusan merupakan hasil dari beberapa proses yang telah dikerjakan serta dilakukan secara berulang-ulang. Partisipan bisa berasal dari suatu organisasi maupun keseluruhan organisasi. Hasil pengambilan keputusan dapat disetujui untuk mencapai suatu tujuan tertentu. Pengambilan keputusan sebenarnya adalah pilihan alternatif dari beberapa solusi. Diharapkan hasil pengambilan keputusan dapat menghasilkan solusi terbaik.

### C. Algoritma dan Metode Klasifikasi

Proses ditemukannya model terbaik yang dapat membantu memvisualisasikan suatu konsep atau kelas sehingga penggambaran dari perbedaan antara dua kelas atau lebih dapat digunakan memprediksi kelas dari data yang belum diketahui termasuk ke kelas yang mana [3]. Data mining merupakan penggalian data sehingga diperoleh informasi tertentu dari dalam suatu database. Klasifikasi adalah bagian dari Data Mining. Selain itu Data Mining menerapkan teknik statistik, matematika, artificial intelligent untuk mendeskripsikan, mengidentifikasi dan menganalisa informasi yang diperoleh dari dari suatu database.

Proses klasifikasi sendiri terdiri dari beberapa empat bagian [2], di antaranya: (1) Kelas yaitu variabel dependen yang berupa kategorikal yang merepresentasikan ‘label’ yang terdapat pada objek. Setelah dilakukan proses data mining maka setiap data yang masuk ke dalam sistem pendukung keputusan akan diberi label/kelas sesuai dengan kondisi/persyaratannya. Contohnya: kategori pendapatan, kategori cuaca, kategori kebangkrutan. (2) Predictor merupakan variabel independen yang digambarkan dengan oleh atribut data. Contohnya: merokok, minum alkohol,

tekanan darah, tabungan, aset, gaji. (3) Data Training merupakan satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor. Jumlah dari data training lebih banyak daripada data testing. Hal ini dimungkinkan karena semakin banyak dan semakin beragam data training yang digunakan maka semakin baik pula akurasi dari sistem pendukung keputusan yang dibangun. (4) Data Testing berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi hasil klasifikasi akan dievaluasi.

#### D. Algoritma K-Nearest Neighbours

KNearest Neighbours merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat dengan titik objek tersebut. Data training diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Nilai k yang terbaik untuk algoritma ini tergantung pada data, secara umumnya, nilai k yang tinggi akan mengurangi efek noise pada klasifikasi, tetapi membuat batasan antara klasifikasi menjadi lebih kabur. Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data training), diantaranya euclidean distance dan manhattan distance (*city block distance*), yang paling sering digunakan adalah euclidean distance [1]. Rumus dari perhitungan euclidean distance dapat dilihat pada persamaan 1.

$$d(y,x)^2=(y_1-x_1)^2+(y_2-x_2)^2+\dots+(y_n-x_n)^2 \quad (1)$$

di mana:

$x_1$  dan  $y_1$ : koordinat data awal (kelas),

$x_2, y_2, x_n, y_n$ : koordinat data testing.

### III. METODE PENELITIAN

Pada bagian ini akan dijelaskan bagaimana metode yang digunakan dalam penelitian ini.

#### A. Data

Data penelitian yang digunakan berupa data mentah dari laporan penerima beasiswa selama tiga tahun terakhir. Data tersebut akan digunakan sebagai data training, sedangkan data testing diperoleh dari data calon penerima beasiswa. Pengumpulan data dalam penelitian ini melalui wawancara dengan staf kemahasiswaan. Jumlah data yang digunakan dalam penelitian ini sebanyak 60 obyek data. Data tersebut terdiri dari data training sebanyak 40 data dan data testing sebanyak 20 data. Obyek data tersebut merupakan data mahasiswa Program Studi Sistem Informasi, Fakultas Vokasi. Data yang diambil adalah data pendapatan orang tua dan Indeks Prestasi Kumulatif (IPK). Data tersebut akan diolah menggunakan Algoritma K-Nearest Neighbours untuk diperoleh klasifikasinya.

#### B. Desain dan Perancangan Sistem

Desain sistemnya akan dibagi menjadi dua desain yaitu desain untuk data training dan data testing. Desain sistem data training ditunjukkan oleh Gbr. 1 sedangkan desain sistem data testing ditunjukkan oleh Gbr. 2.

Variabel yang digunakan dalam pengolahan data adalah variabel Indeks Prestasi Kumulatif (IPK) dan penghasilan orang tua. Penentuan variabel ditentukan dari persyaratan penyeleksian beasiswa. Penentuan variabel dan nilai variabel dapat dilihat pada Tabel 1. Desain antarmuka yang digunakan oleh pengguna dapat dilihat pada Gbr. 4.

TABEL I  
 PENENTUAN VARIABEL DAN NILAI VARIABEL

No	Variabel		Nilai Variabel
1	Indeks Prestasi Kumulatif (IPK)	1.	2,50 – 3,00
		2.	3,01 – 3,50
		3.	3,51 – 4,00
2	Penghasilan Orang Tua	1.	Rp 100,000-Rp 1,000,000
		2.	Rp 1,000,001- Rp 2,000,000
		3.	Rp 2,000,001- Rp 3,000,000
		4.	>Rp 3,000,001

Berdasarkan hasil analisa kebutuhan maka informasi dan kebutuhan yang dibutuhkan adalah perancangan sistem untuk memprediksi penerima beasiswa menggunakan algoritma KNearest Neighbours dalam melakukan prediksi dan klasifikasi. Perancangan sistem yang dilakukan antara lain membuat use case diagram, data flow diagram, activity diagram, sequence diagram dan class diagram.

Diagram alir yang menggambarkan Algoritma KNearest Neighbours pada kasus penentuan penerima beasiswa studi kasus Program Studi Sistem Informasi Fakultas Vokasi digambarkan melalui Gbr. 3.

#### C. Implementasi System

Aplikasi prediksi penerima beasiswa menggunakan Algoritma KNearest Neighbours diimplementasikan menggunakan bahasa Pemrograman Java berbasis desktop. Sehingga penggunaan aplikasi prediksi bersifat internal oleh Program Studi Sistem Informasi.

#### D. Pengujian dan Evaluasi

Evaluasi sistem dilakukan dengan menghitung nilai dari presisi, recall dan akurasi. Nilai presisi adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban dari sistem sedangkan recall adalah tingkat keberhasilan sistem [7]. Akurasi merupakan nilai terdekat antara prediksi dan nilai aslinya. Nilai Precision dan Recall yang baik adalah yang mendekati nilai 1. Persamaan untuk

menghitung nilai dari keduanya ditunjukkan pada persamaan 2 dan persamaan 3. Sedangkan untuk persamaan menghitung akurasi ditunjukkan pada persamaan 4.

$$\text{Presisi} = \text{TP}/(\text{TP}+\text{FP}) \quad (2)$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN}) \quad (3)$$

$$\text{Akurasi} = (\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN}) \quad (4)$$

di mana:

TP : *True Positive*,

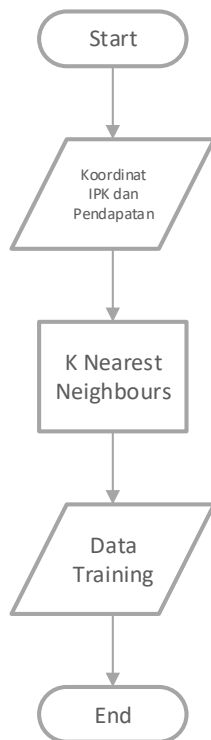
FP : *False Positive*,

FN : *False Negative*,

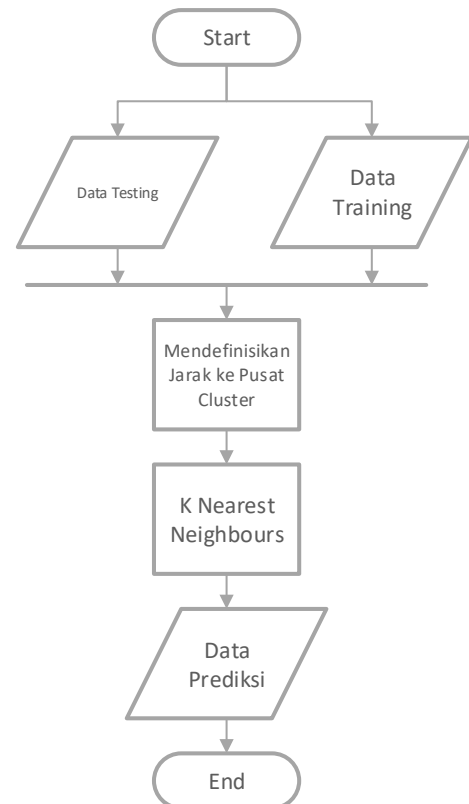
TN : *True Negative*.

#### E. Hasil

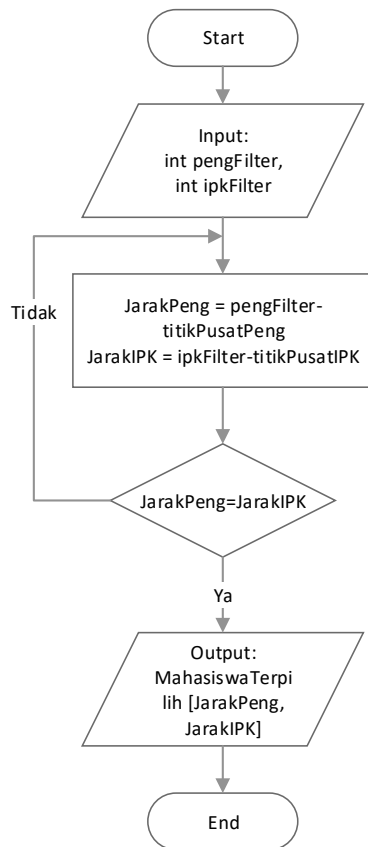
Grup cluster yang dihasilkan ada dua yaitu penerima dan bukan penerima. Dari 20 data testing yang digunakan dalam uji coba, hasil akurasi yang diperoleh adalah 75%. Lima belas data sesuai dengan kategori sedangkan empat data lain tidak sesuai dengan kategori sebenarnya. Tersisa satu data yang tidak bisa dikategorikan karena nilai parameter K yang digunakan adalah 2



Gbr. 1 Diagram Alir Data Training



Gbr. 2 Diagram Alir Data Testing



Gbr. 3 Diagram Alir Algoritma KNearest Neighbours pada Kasus Penentuan Penerima Beasiswa

Gbr. 4 Antarmuka Sistem Prediksi Penentuan Penerima Beasiswa Dengan Metode KNearest Neighbours

### F. Kesimpulan

Evaluasi sistem dari penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa merancang dan membangun sistem pendukung keputusan penerima beasiswa dengan menggunakan Algoritma KNearest Neighbours menghasilkan tingkat akurasi sebesar 75% dengan nilai parameter K yang digunakan adalah 2. Data yang digunakan dalam sistem berjumlah 60 obyek data dengan pembagian 40 data training dan 20 data testing. Saran dari penelitian ini yaitu untuk meningkatkan tingkat keberhasilan sistem pendukung keputusan dapat menambahkan jumlah data mahasiswa dan

perbandingan algoritma yang digunakan pada studi kasus yang sama.

### REFERENSI

- [1] Bramer, M., 2007. Principles of Data Mining, Springer, London.
- [2] Gorunescu, F., 2011. Data Mining : Concepts, Models, and Techniques. Verlag Berlin Heidelberg, Springer.
- [3] Han, J., Kamber, M., 2006. Data Mining Concept and Tehniques, San Fransisco, Morgan Kauffman.
- [4] Kusriani dan Lutfi, E.T., 2009. Algoritma Data Mining, Andi Publishing, Yogyakarta.
- [5] Surbakti, I., 2002. Sistem Pendukung Keputusan, Institut Teknologi Sepuluh November, Surabaya.
- [6] Turban, E., 2005. Decicion Support System and intelligent System, Andi Publisher. Yogyakarta